International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 7
July 2017

Academic Science

A Novel Method for Lung Nodule Classification

Remya M. M.
Department of Electronics & Communication Engineering, Mar Baselios College of Engineering & Technology, Kerala, India

Dr. M. J. Jayashree, Associate Professor
Department of Electronics & Communication Engineering, Mar Baselios College of Engineering & Technology, Kerala, India

Abstract—Lung cancer is a dangerous disease in our daily life. Approximately 20% of lung nodules lead to cancer. Therefore, it is very necessary to find out the nodule in the early stage itself. A new method is proposed for lung nodule classification based on sparse representation classifier. This method is mainly includes patch based division, feature extraction, best feature selection and finally sparse based classification. A strong feature extraction method is used based on SIFT, HOG and MR8+LBP. The best features are selected using t test. This will reduce the overall time consumption. And the use of sparse classifier increases the classification accuracy compared with the other classifier. The proposed scheme is evaluated using a dataset that are collected from the hospital.

Keywords—sparse representation, t test, patch division, SIFT, feature extraction.

I. INTRODUCTION

We know that lung is an essential organ that performs vital role in every instant of our lives. Lung cancer is a sickness which is characterized by uninhibited growth of cells in the lung. It is the main reason of cancer associated deaths in human being. Lung nodules means, small masses that is present in the lung. They are generally spherical shape. Though, the shape can be changed due to the neighboring anatomical structure like blood vessels, pleural surfaces etc. Intra parenchymal lung nodules are very harmful than those associated with the adjacent anatomical structures[1].

According to their relative locations the lung nodules can be divided into 4 types they are,

1. well-circumscribed (W)
2. vascularized (V)
3. juxta-pleural (J) and
4. pleural-tail (P)

Well-circumscribed (W)- the nodule situated at the centre of the lung and no link with vessels, vascularized (V)- here also the nodule is located at the centre of the lung although it is directly attached to the neighboring vessels, juxta-pleural (J)- a large portion of the nodule is attached with the pleural surface, and pleural-tail (P)- the nodule present near the pleural surface and is coupled by a small tail[1][5] as shown in fig. 1. The survival rates depend upon the early detection.

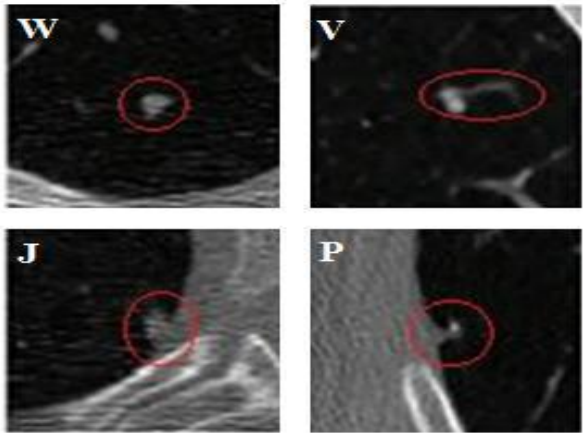

Fig.1: Different types of lung nodule

The treatment of lung cancer is depends upon the type of nodule and nature of nodule that is, whether it is harmful or not. Thus early detection of lung nodule is very essential for increasing the survival rate.

A. Related Works

A lot of studies are reported in the field of lung nodule detection. The most precise imaging technique to get the data for the nodule and its neighboring anatomical structure is computed tomography (CT). However, the analysis of CT image is a difficult process for radiologists because of the huge number of patients. The computer aided diagnosis is more helpful for precise detection. It is based on feature extraction and classification. The

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 7
July 2017

performance of a classifier is depends upon the features. There are many classifier used in the nodule detection field such as support vector machine, decision tree classifier, artificial neural network etc. but they have some limitations.

Some of the earlier works related to lung nodule classification include overlapping nodule identification procedure, which was designed to aid lung nodule classification, but had various defects[3].

II. PROPOSED METHOD

In this paper an efficient method for classification of different types of lung nodule is proposed. The main steps included in this work are a patch based division with concentric multilevel partition, then feature extraction using SIFT(Scale Invariant Feature Transform), HOG (Histogram Of Oriented Gradient) and MR8 (Maximum Response8)+LBP (Local Binary Pattern), best feature selection using t test and sparse representation based classifier to classify the nodule into one of the four types.

The first step is the patch based division. Patch based division methods generally partition the image into circular or square patches. The shape and size of every patch is predetermined. That is, in a rigid manner. Here the partitioning of image is done in an adaptive way. For that an enhanced super pixel segmentation system based on quick shift is used to create patches to construct level nodule and level context. Level nodule means the patch which contains the nodule and level context means the patch which contains the surrounding anatomical structures. The fig. 2 shows the basic block diagram of the proposed scheme.

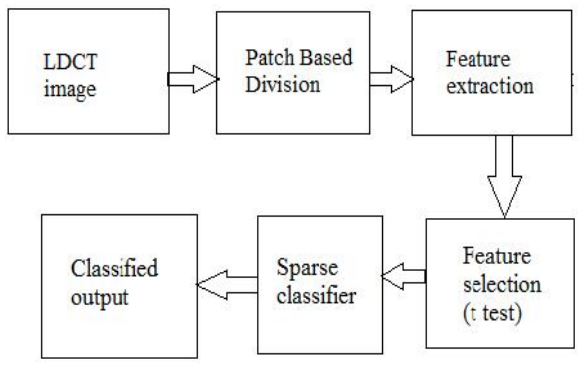

Fig. 2: Block diagram of proposed method

The second step includes feature extraction and best feature selection based on t test. Conventional feature extraction techniques provide large advantages, however highly developed computer vision techniques are now included into medical imaging study. Some of the feature extraction method are, scale-invariant feature transform (SIFT) which is invariant to image translation, rotation and illumination change; local binary pattern (LBP), which gives the information about the texture description of images by using multi scale and rotation invariant feature; histogram of oriented gradient (HOG), which calculates the occurrence of gradient orientation in the local portion of an image[1]. From the large number of extracted features best features are selected using t test, also known as student t test. Thus the overall computation time required for the process become reduces to a very low range.

After the feature selection stage a classifier is required to assign the features for nodule classification. There are many classification methods exist in the field of nodule detection. But accuracy and more time consumption are the most challenging problems. Among these classification methods, sparse representation shows a better and accurate classification result. The percentage of accuracy is more than that of other methods like using SVM.

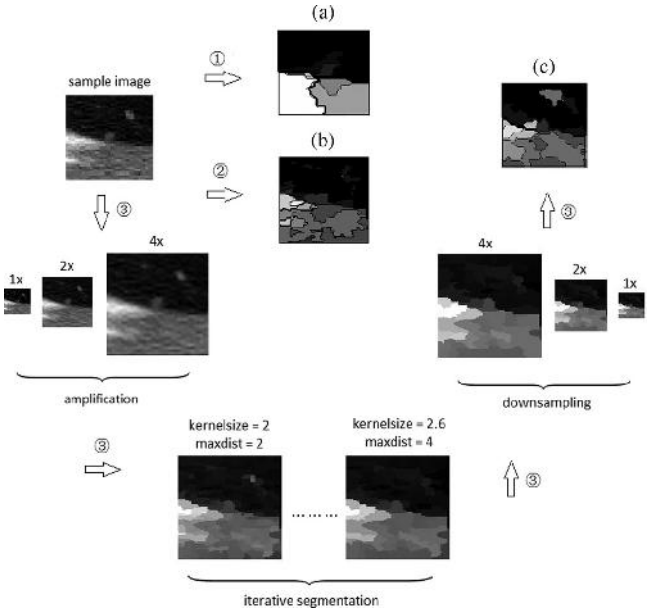

Fig. 3. Formulation of superpixel. (1) direct use of quick shift (2) quick shift based on amplification and downsampling without iterative formulation, (3)

proposed method. Under formulation (shown in (a)); over formulation (shown in (b)), desirable super pixel formulation (shown in (c)).

A. PATCH BASED DIVISION

Usually the patch division methods divide the input image into circular sector or square shapes. That is the division is done in a inflexible manner. Here the size and shape of the patches are predefined. But this will groups the dissimilar objects together. More specifically a fixed partition method may mix various anatomical structures in one patch. This creates difficulty in identification of different types of nodules[1][4].

Normally, the pixels belonging to same patch may contain related information such as gradient and intensities. Therefore an improved patch based partition scheme is used for calculating superpixels based on a quick shift clustering technique. After that, constructs a concentric level partition model by using the distance from the patches to the centroid of the nodule. As a replacement for rigid partitioning scheme here the patches are constructed adaptively using the intensity variation.

a) Superpixel formulation

Formulation of superpixel means the process which dividing an image into many segments. It should contains the narrow information about the image and also reduce the fake labeling due to the noise present in the image. More specifically quick shift clustering method, a mode seeking algorithm is used for forming superpixels from the input image. However, this algorithm cannot be applied directly into an image if the image size is very small. Because, it will gives poor partition result. To overcome this problem quick shift is incorporated in an iterative manner by using amplification and downsampling[4].

The main parameters that are included in quick shift algorithm are kernel size and maximum distance. Kernel size means the size of the kernel which is used to calculate the density. And maximum distance means the distance among the points in the feature space. After iteration process downsampling is used to restore the image into its original size[1].

b) Concentric level partition

In this step depending upon the distance from the patches to the centroid of the nodule patch we divide the patches from one image into multiple concentric levels. To reduce the classification complexity here divides the different levels into two categories such as level nodule and level context. Level nodule gives the information about the nodule present in the input image and level context will gives the surrounding anatomical structure. Sample for concentric level for each kind of lung nodule is shown in fig. 4.

Usually nodules have high contrast than anatomical structures. So we can identify the nodule and anatomical structures. However, the level nodule can include the whole nodule part only if the nodule is isolated from other anatomical structures. And also level nodule might include the surrounding structures if the nodule is analogous to nearby structures. It will create complexity in classification. In these situations over segmentation property of the quick shift method can be in use. Which can be describes the nodule patch based on the middle area of the nodule.

B. EXTRACTION OF FEATURES

The next step is the feature extraction. For an image detection scheme feature extraction stage is the key step. Because based on these features the classifier determines the various kinds of nodule. A good feature descriptor must confine the individual uniqueness and should be flexible to diverse image circumstances. There are many feature descriptors are available. Each one has its advantages. And is depends upon the type of field used. Here a feature set of the mixture of SIFT, MR8+LBP and HOG are used. SIFT will gives the overall description about intensity gradient and texture ,MR8+LBP will provides the information about texture and HOG gives the information about gradient of an image[10].

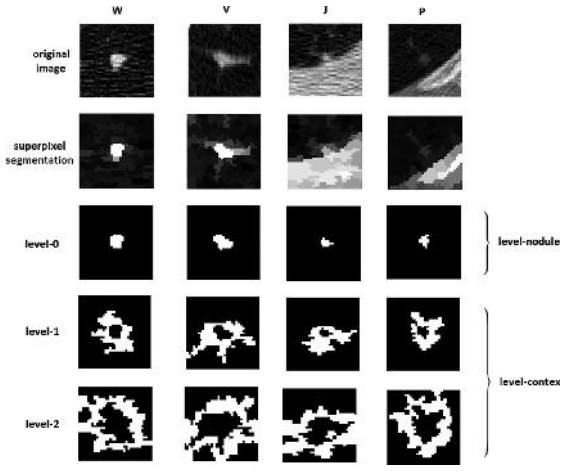

Fig. 4. Concentric level partition for different types.

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 7
July 2017

Academic Science

SIFT will extracts a 128 length vector. It is invariant to image rotation translation and variations in the illumination. The shape of the patches are not same so selected small rectangular window to include all the pixels for the entire patch and ran a SIFT window above the image. MR8 filter bank contains 38 filters. But only 8 responses is provided. It records only the maximum response across all the orientations. LBP is a powerful feature descriptor widely used in classification based on texture of an image. It will capture many minute image variations. Therefore it is used along with the MR8 filter[9]. HOG is widely used in the field of image classification. But it will not handle the rotation invariant problem. It is more helpful to identify the anatomical structures present in the input image. The three feature descriptors will provide a number of features. Then, among these features best features are selected using a test called t test.

a) T test

T test is also known as student t test. It is a statistical hypothesis test. In this paper it is used for best feature selection. From the feature extraction stage we get a large number of features. From that best features will selected using t test. It uses a student t distribution. It can be used if the test statistics allows a normal distribution. And also its scaling term value will be known. If the scaling value is not known then we can substituted it with an estimate value according to the data.

Many types of t test are there. Among that chi square test is the most universal type. It calculates the ratio of difference in mean between two classes. It can used only when the mean of two classes are statistically different. But these are applicable only for two class issues. In the case of more than two class problem we want to calculate the t statistic value for every feature in every class.

$$tic = (xic-xi) / (Mc.(Si+S0))$$

$$Si2 = 1/ (N-C) \Sigma \Sigma (xij-xic)^2$$

$$Mc = \sqrt{(1/nc + 1/N)}$$

where tic - i-th feature's t-statistics value for c-th class

xic - mean of the i-th feature in the c-th class,

xi - mean of i-th feature for all classes

xij - i-th feature of the j-th sample;

N - Number of samples in class C and

nc - the number of samples in class c;

Si - standard deviation and

S0 is set as the median value of Si for all the features.

$$t_i = max\left\{\frac{|\overline{x}_{ic} - \overline{x}_i|}{M_c S_i}, c = 1, 2, ...C\right\}$$

It can be generalized as;

1. Let the feature set is defined as F = {fi,…fj…fg}, and i-th feature has diverse m_i values denoted as fi= {xi1,xi2....xim}.

2. Change every actual feature value into a vector by dimension mi:

$$x_i^{(1)} \Rightarrow X_i^{(1)}=(0, \ldots, 0, 1), x_i^{(2)} \Rightarrow X_i^{(2)}=(0, \ldots, 1, 0), \ldots, x_i^{(m_i)} \Rightarrow \vec{X}_i^{(m_i)}=(1, \ldots, 0, 0).$$

3. Substitute this vector for all the numerical features.

$$t_i = max\left\{\frac{\left|\overline{\vec{X}}_{ic} - \overline{\vec{X}}_i\right|}{M_c S_i}, c = 1, 2, ...C\right\}$$

$$S_i^2 = \frac{1}{N - C}\sum_{c=1}^{C}\sum_{j \in c}(\vec{X}_{ij} - \overline{\vec{X}}_{ic})(\vec{X}_{ij} - \overline{\vec{X}}_{ic})^T$$

C. CLASSIFICATION

After feature extraction stage the next step is classification. Classification is the process of find out the type nodule present in the input image based on the extracted features. Many classifiers are available in the field of nodule detection. But each of them has some limitations. In this paper sparse classifier is used for Lung Nodule Classification. Sparse representation based classifier are widely used in many field such as signal processing image processing etc. because of its simplicity and flexibility. Sparse classifiers can achieve high detection sensitivity and accuracy than other methods. Sparseness means scattering of piecewise component. It is an enviable feature in classifier because,

- It leads to structural simplification
- Performance can increases with degree of sparseness

Consider the equation,

$$W = \beta X$$

Where β is the sparse component and W is the test sample image. and X denotes the input training features and is defined as X = [X_1, X_2….Xn]. We have to choose a β value that minimizes the error.

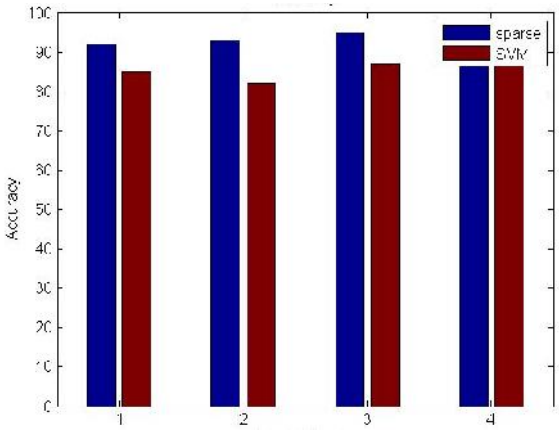

Fig. 5. Comparison of classification rate of proposed method with SVM for W, V, J, and P types respectively.

III. RESULTS AND DISCUSSION

Adaptive patch based division using improved quick shift algorithm Concentric level partition can achieve better partitioning of the image. This iterative scheme outperformed than other methods. And also the size and shape of the patches are not predefined. The use of features descriptors such as SIFT, HOG, and MR8+LBP can provide better performances than other methods. This improves the classification accuracy. In the conventional methods the number of features that are used for classification is very large. That will increase the computation time. Therefore, the time required for overall process is very much high. But in this proposed method t test is used for feature selection. It selects the best features based on a probability value. Thus the numbers of features that are used for classification are very low. Therefore it is understandable that the time required for proposed scheme is very less than that of conventional methods. The use of sparse classifier provided greater classification accuracy than the earlier methods like SVM. Thus the efficiency of the system is greatly enhanced by the use of sparse classifier in the system. The performance of proposed scheme is compared with the SVM classifier and is shown in fig. 5. From the graph it is understandable that the performance of this scheme increased.

IV. CONCLUSION

This paper presents an efficient classification scheme for lung nodule LDCT images. Lung plays an important role in human life. Lung cancer is a sickness which is described by uninhibited cell expansion in tissues of the human lung. Around 20% of the lung tumor leads to cancer. Hence it is significant to recognize whether it is harmful or not. There are many methods to find out the nodules in the lung. But they are less efficient if the image is too small or the image is a low dose CT scan. In this project a patch based division based on improved quick shift clustering algorithm is used, which is more favorable for small images as well as LDCT images. Then, three feature descriptor containing SIFT, MR8+LBP, and HOG are used to illustrate the image patch features. Then among these features only best features are selected for the classification purpose by using a method called t test. Therefore the overall time required for the process becomes very less compared with the conventional methods. Finally, a sparse representation based classifier is designed to find out which type of nodule is present in the image. The classification rate of sparse classifier is more than that of other classifier which improves the performance of the system.

References

[1] Fan Zhang ,Yang Song, Weidong Cai, Min-Zhao Lee, Yun Zhou, Heng Huang,Shimin Shan, Fulham M.J., Feng D.D., (2014), " Lung Nodule Classification With Multilevel Patch Based Context Analysis", IEEE transaction on Biomedical Engineering,volume:61, Page(s): 1155 – 1166.

[2] M H Hasna, Jobin Jose , 2015 "Lung nodule classification using multilevel patch-based context analysis and decision tree classifier" IJERT,vol. 02, 2290-2294.

[3] Md. Badrul Alam Miah, Mohammad Abu Yousuf "Detection of Lung Cancer from CT Image Using Image Processing and Neural Network" International Conference, 21-23 May 2015.

[4] Abina Aliyar, Veena Vijayan, (2015) " Adaptive Multilevel Patch Based Lung Nodule Classification" IJESC, ISSN-2321 -3361.

[5] A. A. Farag, "A variational approach for small-size lung nodule segmentation," in Proc. Int. Symp. Biomed. Imag., 2013, pp. 81–84.

[6] L. Fan, C. L. Novak, J. Qian, G. Kohl, and D. Naidich, "Automatic detection of lung nodules from multislice low-dose CT images," in Proc. SPIE, Med. Imag., 2001, vol. 4322, pp. 1828- 1835.

[7] A. Farag, S. Elhabian, J. Graham, A. Farag, and R. Falk, "Toward precise pulmonary nodule descriptors for nodule type classification," in Proc. Med. Image Comput. Comput.-Assisted Intervention Conf. Lecture Notes Comput. Sci., 2010, vol. 13, no. 3, pp. 626–633.

[8] S. L. A. Lee, A. Z. Kouzani, and E. J. Hu, "Automated detection of lung nodules in computed tomography images: A review," *Mach. Vis. Appl.*, vol. 23, no. 1, pp. 151–163, 2012.

[9] Y. Song, W. Cai, Y. Wang, and D. Feng, "Location classification of lung nodules with optimized graph construction," in *Proc. Int. Symp. Biomed. Imag.*, 2012, pp. 1439–1442.

[10] A. Farag, A. Ali, J. Graham, S. Elshazly, and R. Falk, "Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest," in *Proc. Int. Symp. Biomed. Imag.*, 2011, pp. 169–172.